# Supplementary Materials for:

# Amino Acid Position 11 of HLA-DRβ1 is a Major Determinant of Chromosome 6p Association with Ulcerative Colitis

Jean-Paul Achkar[1,2]*, Lambertus Klei[3]*, Paul I.W. de Bakker[4-6,]*, Gaia Bellone[7], Nancy Rebert[2], Regan Scott[8], Ying Lu[9], Miguel Regueiro[8], Aaron Brzezinski[1], M. Ilyas Kamboh[10], Claudio Fiocchi[1,2], Bernie Devlin[3,10], Massimo Trucco[9], Steven Ringquist[9], Kathryn Roeder[7,11#], Richard H Duerr[8,10#]

[1]Department of Gastroenterology & Hepatology, Digestive Disease Institute, Cleveland Clinic, Cleveland, Ohio

[2]Department of Pathobiology, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio

[3]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

[4]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

[5]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts

[6]Julius Center for Health Sciences and Primary Care, and Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

[7]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania

[8]Division of Gastroenterology, Hepatology, & Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

[9]Division of Immunogenetics, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, Pennsylvania

[10]Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania

[11]Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania

*These authors contributed equally to this work

#These authors contributed equally to this work

# Supplementary Methods

### *Assessment of HLA-DRB1 imputation quality*

*HLA-DRB1* imputation quality at two-digit resolution was assessed from 384 of our study subjects using sequence-specific oligonucleotide probes (SSO) provided by the LABType SSO Typing Tests (One Lambda, Canoga Park, CA) as well as by next-generation sequencing of polymerase chain reaction (PCR) amplified DNA using the Roche/454 titanium assay. SSO typing employed specific oligonucleotide probes immobilized on fluorescently coded microspheres. A LABScan 100 flow analyzer (Luminex, Austin, TX) was used to quantify fluorescent intensity of phycoerythrin (PE) resulting from hybridization of PE-streptavidin labeled biotinylated amplified DNA. Briefly, PCR product was generated from 40ng genomic DNA using primers, buffers, dNTPs, and Taq polymerase provided by the manufacturer. The reaction mixture was incubated at 96°C for 3 minutes followed by 5 thermal cycles at 96°C for 20 seconds, 60°C for 20 seconds, and 72°C for 20 seconds. A second series of 30 thermal cycles was performed at 96°C for 10 seconds, 60°C for 15 seconds, and 72°C for 20 seconds followed by incubation at 72°C for 10 minutes. Amplified DNA was denatured using Denaturation Buffer (One Lambda), the pH was neutralized using Neutralization Buffer (One Lambda), and allowed to hybridize with complementary DNA probes conjugated to fluorescently coded microspheres at 60°C for 15 minutes. Nonspecifically bound DNA was removed by 3 washes using One Lambda Wash Buffer. Labeling was performed by incubation with PE labeled streptavidin at 60°C for 5 minutes. Unbound streptavidin was removed by washing with One Lambda Wash Buffer. Samples were scanned for

PE fluorescent intensity using a LABScan 100 flow analyzer and the acquired data analyzed using LABType SSO Analysis Software purchased from One Lambda.

Next generation sequencing assays and analysis software for genotyping *HLA-DRB1* were developed in house. Next-generation sequences were obtained using the Roche/454 titanium assay and were performed at the University of Pittsburgh Genomics and Proteomics Core Laboratories. Sequence (.fna) and quality-score (.qual) data files were generated using Roche Genome Sequencer software. HLA genotyping was performed using the CAPSeq analysis software (Gaia Bellone, unpublished data) which incorporated MUMmer[1] to align sequences and the R package diffusionMap[2] to cluster the sequence data into allelic groups. Genotypes were determined by comparing the consensus sequences obtained from each cluster with sequences of known HLA-DRB1 alleles obtained from the IMGT/HLA Database.[3]

Briefly, barcoded PCR oligonucleotide primers (forward primer 5'-CCGGATCCTTCGTGTCCCCACAGCACG-3', reverse primer 5'-CCGCTGCACTGTGAAGCTCTC-3') were used to target *HLA-DRB1* exon 2 for amplification.[4] DNA amplification were performed in 10 ul and used 40ng genomic DNA, 400nM oligonucleotide primers, 0.4mM dNTP, and 1 unit FastStart High Fidelity Enzyme Blend (Roche Diagnostics). Thermal cycling was performed by incubation at 94°C for 3 minutes followed by 39 cycles at 94°C for 30 seconds, 61°C for 45 seconds, and 72°C for 1 minute. The reactions were then incubated at 72°C for 2 minutes. Following small DNA fragments removal via AMPure beads (Agencourt, Danvers, MA) a 1 ul aliquot was run on a bioanalyzer DNA 7500 chip (Agilent, Santa Clara, CA) to verify size distribution.

Next, emulsion PCR (emPCR) was performed using single stranded DNA fragments bound to Library capture Beads by mixing sample library with capture beads and performing the following thermocycler program: 80°C for 5 minutes, ramp 0.1°C /sec to 70°C, hold at 70°C for 1 minute, ramp 0.1°C/sec to 60°C, hold at 60°C for 1 minute, ramp 0.1°C/sec to 50°C, hold at 50°C for 1 minute, ramp 0.1°C/sec to 20°C. Emulsion of PCR reagents in microreactors with library capture beads was prepared by mixing beads, PCR reaction mix (1X amplification mix, Amplification Primers, 0.15 U/ul Platinum Taq (Invitrogen, Carlsbad, CA)), and emulsion oil and mixing vigorously using a Tissue Lyser (Qiagen, Valencia, CA). Emulsion was distributed into a PCR plate and template amplification was carried out in a thermocycler using the following cycling conditions: Hotstart activation for 4 minutes at 94°C, 40 cycles of 94°C for 30 seconds, 58°C for 1 minute, 68°C for 90 seconds followed by 13 cycles of 94°C for 30 seconds and 58°C for 6 minutes. Following template amplification, emulsions were broken and beads with amplified product recovered by repeated washes with ethanol using a syringe filter unit.

Bead enrichment was accomplished using enrichment beads added to the recovered amplification beads. Enrichment beads were coated with oligonucleotides complementary to the free end of the amplified template. Successful amplification beads were bound to the paramagnetic enrichment beads and drawn out of solution using a magnetic rack. Unsuccessful amplification beads were drawn off with the supernatant and discarded. The bond between the amplification and enrichment beads was broken using 125 mM NaOH. Enrichment beads were pelleted using a magnetic rack and the enriched amplification beads were recovered. Melt solution was

neutralized by repeated washes with 1X Annealing buffer and the beads were left suspended in annealing buffer.

Sequencing primers were added to the mixture of beads and annealing buffer and annealed to the template using the following thermocycler conditions: 65°C for 5 minutes, ramp to 50°C at 0.1°C/ second, hold at 50°C for 1 minute, ramp to 40°C at 0.1°C/ second, hold at 40°C for 1 minute, ramp to 15°C at 0.1°C/second, hold 15°C. Beads were counted using a Beckman Z1 particle counter. Picotiter plates were prepared based on bead count obtained above and manufacturer recommendations for the picotiter plate region size being used, control beads and sample beads are mixed to form the sequencing sample. Packing beads, sample beads and enzyme beads were applied to the picotiter plate as per manufacturer instructions. The sequencing reaction was performed in picotiter plates loaded onto the FLX sequencer and the run started. The FLX used pyrosequencing chemistry and detected the incorporation of each nucleotide in real time.

Emulsion PCR Titration was performed prior to next-generation sequencing. For each sample, a preliminary titration run was performed in order to determine the best ratio of DNA template to amplification beads to obtain a maximum amount of usable sequence from the final data run. The concentration of DNA template was determined in copies/ul by applying the mass of the average fragment size as determined from the bioanalyzer output to the measured concentration of amplicon DNA library. emPCR reactions corresponding to titration points of 0.5, 2, 4 and 16 copies/bead were performed as described for emPCR above. Emulsions were broken and beads recovered. Amplification beads were not enriched but were counted and loaded onto a

picotiter plate for sequencing analysis. The number of beads that could be successfully sequenced was used to determine the optimum ratio of DNA to beads for the final emPCR reaction to produce template for data production.

## Supplementary Results

### *Concordance of HLA-DRB1 typing by imputation, SSO and sequencing*

We found concordance between *HLA-DRB1* imputation results and SSO results for 374 out of 383 samples (one sample failed to genotype by SSO). In the 9 discrepant cases, the discordance resulted from mismatched genotypes at only one of the two alleles and thus when we analyzed the total of 766 alleles, the concordance rate was 98.8% between the two methods (Supplementary Table 1). Confirmation of this conclusion was obtained using next-generation sequence based typing where we found concordance for 368 out of 384 samples. In the 16 discrepant cases observed between imputation and sequencing the differences were due to the same 9 mismatched genotypes identified previously by SSO plus an additional 7 mismatched genotypes (Supplementary Table 1). It was notable that in the latter comparison all 7 mismatches were consistent with allele dropout in the sequencing results from unequal PCR amplification.[5, 6] Therefore, discordant genotypes occurred for only one of the two alleles and when we analyzed the total of 768 alleles, the concordance rate between imputation and sequencing was 97.9%. In contrast, the concordance rate between SSO and next-generation sequencing was 99.1% (759 out of 766 alleles) resulting solely from the occurrence of 7 homozygous genotypes called by the latter method. The uncertainty that occurred during imputation of *HLA-DRB1* genotype did not reach a level that significantly impacted our analyses. Comparison of imputation with the results of either SSO or next-generation sequence based genotyping methods indicates that the imputation procedure we applied was highly accurate.

**Supplementary Table 1.** Comparison of discordant results for imputed *HLA-DRB1* genotypes versus genotypes determined using SSO and sequencing. Alleles in red font indicate discrepant results compared to imputed genotypes. SSO, sequence-specific oligonucleotide probes; NR, no result (sample failed to genotype).

| | Imputed Genotype | | SSO Genotype | | Sequencing Genotype | |
|---|---|---|---|---|---|---|
| Subject | Allele 1 | Allele 2 | Allele 1 | Allele 2 | Allele 1 | Allele 2 |
| A | *DRB1*01* | *DRB1*13* | *DRB1*01* | *DRB1*11* | *DRB1*01* | *DRB1*11* |
| B | *DRB1*01* | *DRB1*13* | *DRB1*01* | *DRB1*11* | *DRB1*01* | *DRB1*11* |
| C | *DRB1*03* | *DRB1*09* | *DRB1*03* | *DRB1*03* | *DRB1*03* | *DRB1*03* |
| D | *DRB1*03* | *DRB1*04* | *DRB1*03* | *DRB1*07* | *DRB1*03* | *DRB1*07* |
| E | *DRB1*11* | *DRB1*12* | *DRB1*08* | *DRB1*11* | *DRB1*08* | *DRB1*11* |
| F | *DRB1*13* | *DRB1*15* | *DRB1*08* | *DRB1*15* | *DRB1*08* | *DRB1*15* |
| G | *DRB1*11* | *DRB1*12* | *DRB1*11* | *DRB1*11* | *DRB1*11* | *DRB1*11* |
| H | *DRB1*13* | *DRB1*15* | *DRB1*11* | *DRB1*15* | *DRB1*11* | *DRB1*15* |
| I | *DRB1*13* | *DRB1*15* | *DRB1*11* | *DRB1*15* | *DRB1*11* | *DRB1*15* |
| J | *DRB1*04* | *DRB1*04* | NR | NR | *DRB1*04* | *DRB1*04* |
| K | *DRB1*03* | *DRB1*13* | *DRB1*03* | *DRB1*13* | *DRB1*03* | *DRB1*03* |
| L | *DRB1*07* | *DRB1*13* | *DRB1*07* | *DRB1*13* | *DRB1*07* | *DRB1*07* |
| M | *DRB1*01* | *DRB1*13* | *DRB1*01* | *DRB1*13* | *DRB1*01* | *DRB1*01* |
| N | *DRB1*11* | *DRB1*16* | *DRB1*11* | *DRB1*16* | *DRB1*16* | *DRB1*16* |
| O | *DRB1*04* | *DRB1*13* | *DRB1*04* | *DRB1*13* | *DRB1*04* | *DRB1*04* |
| P | *DRB1*11* | *DRB1*16* | *DRB1*11* | *DRB1*16* | *DRB1*11* | *DRB1*11* |
| Q | *DRB1*13* | *DRB1*15* | *DRB1*13* | *DRB1*15* | *DRB1*15* | *DRB1*15* |

## Supplementary References


1. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C *et al.* Versatile and open software for comparing large genomes. *Genome Biol* 2004; **5:** R12.

2. Richards J, Freeman P, Lee A, Schafer C. Exploiting low-dimensional structure in astronomical spectra. *Astophysical Journal* 2009; **691:** 32-42.

3. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. *Nucleic Acids Res* 2011; **39:** D1171-6.

4. Bentley G, Higuchi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg EA *et al.* High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens* 2009; **74:** 393-403.

5. Voorter CE, Kik MC, van den Berg-Loonen EM. High-resolution HLA typing for the DQB1 gene by sequence-based typing. *Tissue Antigens* 1998; **51:** 80-7.

6. van Dijk A, Melchers R, Tilanus M, Rozemuller E. HLA-DQB1 sequencing-based typing updated. *Tissue Antigens* 2007; **69** (Suppl 1): 64-5.